DETECTION OF LOCAL VISUAL SPACE-TIME DETAILS IN A VIDEO SIGNAL

FIELD OF THE INVENTION

The present invention relates to the field of video signal processing such as for TV or DVD signals. More specifically, the invention relates to methods for detection and segmentation of local visual space-time details in video signals. In addition, the invention

5   relates to systems for detection and segmentation of local visual space-time details in video signals.

BACKGROUND OF THE INVENTION

Data compression of video signal with a stream of images (frames) has become

10   widespread since a large amount of channel or storage capacity can be saved in transmission of digital video data such as for TV or DVD. Specified standards such as MPEG and H.26x provide a high degree of data compression using block-based motion compensation techniques. Normally, macro-blocks of 16x16 pixels are used for representation of motion information. For many normal video signals these compression techniques provide a high

15   data compression rate without suffering from any visual artefact that is perceptible by the human eye.

However, the standard compression schemes are known not to be transparent, i.e. for certain video signals they give rise to visual artefacts. Such visual artefacts occur in case the

20   video signal includes motion pictures including local space-time details. Local space-time details are represented by spatial texture that varies its local characteristics in time in an indefinite manner. Examples are motion pictures of fire, wavy water, rising steam, leaves fluttering in the wind etc. In these cases the motion picture information representation by 16x16 pixel macro-blocks offered by the compression schemes is too coarse to avoid loss of

25   visual information. This is a problem in relation to achieve optimal high quality video reproduction in combination with the benefits of MPEG or H.26x compression with respect to bit rate reduction.

In order to avoid visual artefacts in a video signal intended for compression, it is

30   necessary to detect local space-time details that may cause visual artefacts by compression prior to applying the compression procedure. Having located these parts in the video signal it is possible to apply a special processing to these parts so as to avoid artefacts being

introduced by the compression procedure. Methods for detecting and indicating image blocks
of a video signal that include space-time details are known.

5   EP 0 571 121 B1 describes an image processing method being an elaboration of the
known so-called Horn-Schunk method. This method is described in B. K. Horn, and B. G.
Schunck, "Determining Optical Flow", Artificial Intelligence, Vol. 17, 1981, pp. 185-204.
The Horn-Schunk method includes extraction of pixel-wise image velocity information called
optical flow. For each single image an optical flow vector is determined, and a condition
number is computed based on this vector. In EP 0 571 121 B1 a local condition number is
10 computed based on the optical flow vector for each image, the goal being to obtain a robust
optical flow.

EP 1 233 373 A1 describes a method for segmentation of fragments of an image
exhibiting similarities in various visual attributes. Various criteria are described for
15 combining small regions of an image into larger regions exhibiting similar characteristics
within a predetermined threshold. In relation to detection of motion an affine motion model is
used which implies calculation of optical flow.

US 6,456,731 B1 describes a method for estimation of optical flow and an image
20 synthesis method. The described estimation of optical flow is based on the known Lucas-
Kanade method described in B. D. Lucas, and T. Kanade, "An iterative image registration
technique with an application to stereo vision", Proceedings of the 7th International Joint
Conference on Artificial Intelligence, 1981, Vancouver, pp. 674-679. The Lucas-Kanade
method estimates optical flow by assuming that optical flow is constant within a local
25 neighbourhood of a pixel. The image synthesis method is based on a process of registering
consecutive images of a sequence by using values of estimated optical flow and a velocity of
specifically tracked image points, visually salient like corner points, using the known so-
called Tomasi-Kanade temporal feature tracking method. Thus, the method described in US
5,456,731 B1 does not perform image partitioning, but similar to the method described in EP
30 0 571 121 B1, it performs the step of computing optical flow, and subsequently the step of
image registering.

SUMMARY OF THE INVENTION
It may be seen as an object of the present invention to provide a method of detecting
35 local space-time details in a video signal. The method must be simple to implement and it
must be adapted for application within low cost equipment. By space-time details of an
image is understood image regions containing a large spatial brightness variation that exhibits

strong temporal changes at the local level, wherein a velocity of these spatial parts are weakly correlated in time.

A first aspect of the present invention provides a method of detecting local space-time details of a video signal representing a plurality of images, the method comprising, for each image, the steps of:

A) dividing the image into one or more blocks of pixels,

B) calculating at least one space-time feature for at least one pixel within each of said one or more blocks,

C) calculating for each of the one or more blocks at least one statistical parameter for each of the at least one space-time features calculated within the block, and

D) detecting blocks wherein the at least one statistical parameter exceeds a predetermined level.

Preferably, the at least one space-time feature comprises visual normal flow magnitude and/or visual normal flow direction. The visual normal flow represents the component of the optical flow that is parallel to image brightness spatial gradient. The at least one space-time feature may further comprise visual normal acceleration magnitude and/or visual normal acceleration direction. Visual normal acceleration describes temporal variation of the visual normal flow along the normal (image brightness gradient) direction.

Preferably, the method further comprises the steps of calculating horizontal and vertical histograms of the at least one space-time feature calculated in step C).

The at least one statistical parameter of step D) may comprise one or more of: variance, average, and at least one parameter of a probability function. The block(s) of pixels are preferably non-overlapping square blocks, and their size may be: 2x2 pixels, 4x4 pixels, 6x6 pixels, 8x8 pixels, 12x12 pixels, or 16x16 pixels.

The method may further comprise the step of pre-processing the image prior to applying step A), so as to reduce noise in the image, this pre-processing preferably comprising the step of convolving the image with a low-pass filter.

The method may further comprise an intermediate step between step C) and D), the intermediate step comprising calculating at least one inter-block statistical parameter involving at least one of the statistical parameter calculated for each block. The at least one inter-block statistical parameter may be calculated using a 2-D Markovian non-causal neighbourhood structure.

The method may further comprise the step of determining a pattern of temporal evolution for each of the at least one statistical parameter calculated in step C). The method may further comprise the step of indexing at least part of an image comprising one or more blocks detected in step D). Furthermore, the method may comprise the step of increasing data rate allocation to the one or more blocks detected in step D). In another embodiment, the method may further comprise the step of inserting an image in a de-interlacing system.

A second aspect of the invention provides a system for detecting local space-time details of a video signal representing a plurality of images, the system comprising:
- means for dividing an image into one or more blocks of pixels,
- space-time feature calculating means for calculating at least one space-time feature for at least one pixel within each of the one or more blocks,
- statistical parameter calculating means for calculating for each of the one or more blocks at least one statistical parameter for each of the at least one space-time features computed within the one or more blocks, and
- detecting means for detecting one or more blocks wherein the at least one statistical parameter exceeds a predetermined level.

A third aspect of the invention provides a device comprising a system according to the system of the second aspect.

A fourth aspect of the invention provides a signal processor system programmed to operate according to the method of the first aspect.

A fifth aspect of the invention provides a de-interlacing system for a television (TV) apparatus, the de-interlacing system operating according to the method of the first aspect.

A sixth aspect provides a video signal encoder for encoding a video signal representing a plurality of images, the video signal encoder comprising:
- means for dividing an image into one or more blocks of pixels,
- space-time feature calculating means for calculating at least one space-time feature for at least one pixel within each of the one or more blocks,
- statistical parameter calculating means for calculating for each of the one or more blocks at least one statistical parameter for each of the at least one space-time features computed within the one or more blocks,
- means for allocating data to the one or more blocks according to a quantisation scale, and
- means for adjusting the quantisation scale for the one or more blocks in accordance with the at least one statistical parameter.

A seventh aspect provides a video signal representing a plurality of images, the video signal comprising information regarding image segments exhibiting space-time details suitable for use with the method of the first aspect.

5   An eighth aspect provides a video storage medium comprising video signal data according to the seventh aspect.

A ninth aspect provides a computer useable medium having a computer readable program code embodied therein, the computer readable program code comprising:

10   - means for causing a computer to read a video signal representing a plurality of images,
       - means for causing the computer to divide a read image into one or more blocks of pixels,
       - means for causing the computer to calculate at least one space-time feature for at least one pixel within each block,
       - means for causing the computer to calculate for each of the blocks at least one statistical

15   parameter for each of the at least one space-time features calculated within the one or more blocks, and
       - means for causing the computer to detect blocks wherein the at least one statistical parameter exceeds a predetermined level.

20   A tenth aspect provides a video signal representing a plurality of images, the video signal being compressed according to a video compression standard, such as MPEG or H.26x, comprising a specified individual allocation of data to blocks of each image, wherein a data rate allocated to one or more selected blocks of images exhibiting space-time details is increased compared to the specified allocation of data to the one or more selected blocks.

25
       An eleventh aspect provides a method of processing a video signal, wherein the method of processing comprises the method of the first aspect.

       A twelfth aspect provides an integrated circuit comprising means for processing a

30   video signal according to the method of the first aspect.

       A thirteenth aspect provides a program storage device readable by a machine and encoding a program of instructions for executing the method of the first aspect.

35   **BRIEF DESCRIPTION OF DRAWINGS**
       In the following the invention is described in details with reference to the accompanying figures, wherein

Fig. 1 shows an illustration of normal and tangential flows at two points of a contour moving with uniform velocity,

Fig. 2a shows an example of an image of two persons and a fountain basin including

5   splashing water,

Fig. 2b shows a grey scale plot representing for the image of Fig. 2a a block-wise level of normal flow variance, wherein white blocks indicate blocks calculated to have a high level of normal flow variance,

10

Fig. 3 shows a flow diagram of a system according to the present invention, and

Fig. 4 shows an example of a normal flow variance histogram.

15   While the invention is susceptible to various modifications and alternative forms, specific embodiments have been shown by way of example in the drawings and will be described in detail herein. It should be understood, however, that the invention is not intended to be limited to the particular forms disclosed. Rather, the invention is to cover all modifications, equivalents, and alternatives falling within the scope of the invention as

20   defined by the appended claims.

DETAILED DESCRIPTION OF THE INVENTION

According to an embodiment of the present invention the major operations to be carried out for processing an image are the steps:

25

A) Divide image into blocks
B) Estimate local feature(s)
C) Calculate feature statistics per block

30   Step A) of processing an image is to divide the image into blocks. Preferably, the blocks coincide with macro blocks used by standard compression such as MPEG and H.26x. Therefore, the image is preferably divided into non-overlapping blocks of 8x8 pixels or 16x16 pixels. The image blocks, when 8x8 pixels large and when they are aligned with the (MPEG) image grid, coincide with typical I-frame DCT/IDCT computation and describe

35   spatial details information. When 16x16 pixels large and when they are aligned with the (MPEG) image grid, coincide with P-frame (B-frame) macro blocks for doing motion compensation (MC) in block-based motion estimation in MPEG/H.26x video standards, and this allows to describe spatio-temporal details information.

Step B) comprises estimating at least one local feature, the local feature relating to spatial, temporal, and/or spatio-temporal details of the image. Preferably, two features are used together with different associated metrics. The estimation of local features is based on a combination of spatial and temporal image brightness gradients. The preferred features are

5   visual normal flow, i.e. and visual normal velocity and visual normal acceleration. The local feature may be based on either or both of visual normal velocity and visual normal acceleration. For the case of visual normal velocity two consecutive frames (or images) are used, while for the visual normal acceleration three consecutive frames (or images) are necessary. A more thorough description of visual normal velocity and visual normal

10   acceleration is given in the following.

Step C) comprises calculating a per block feature statistics. This includes the computation of feature average and variance. Also, different probability density functions are matched to this per block statistics. The per block statistics provides information so as to set

15   up thresholds or criteria allowing a categorisation of each block with respect to the amount of space-time details. Thus, the per block statistics allows detection of blocks with a high amount of space-time details, since such blocks exhibit per blocks statistical parameters exceeding predetermined thresholds.

20   The visual normal flow represents the component of the optical flow that is parallel to image brightness spatial gradient. Optical flow is the most detailed velocity information that can be extracted locally by processing two successive frames or video fields, but it is computationally expensive to extract. The normal flow, on the other side, is easy to compute and it is very rich in local spatial and temporal information. For example calculation of

25   optical flow requires typically 7x7x2 space-time neighbourhoods, while normal flow requires only 2x2x2 neighbourhoods. In addition, calculation of optical flow requires an optimisation, while calculation of normal flow does not.

The normal flow magnitude determines the amount of motion parallel to the local

30   image brightness gradient and the normal flow direction describes the local image brightness orientation. Visual normal flow is calculated from:

$$v_x \times \frac{\partial I(x,y,t)}{\partial x} + v_y \times \frac{\partial I(x,y,t)}{\partial y} + \frac{\partial I(x,y,t)}{\partial t} = 0 \,,$$

where I is brightness, x and y are spatial variables, and t is the time variable. The normal flow

35   direction encodes implicitly spatial variation of image brightness gradient and therefore spatial texture information. The normal acceleration describes, as a second order effect, how the normal flow varies locally.

Visual normal flow is defined as the normal, i.e. parallel to the spatial image gradient, component of the local image velocity or optical flow. The image velocity can be decomposed, at each image pixel, into normal and tangential components.

Fig. 1 shows, for illustration, a well-defined image boundary or contour that passes the target pixel of an image. The diagram in Fig. 1 shows the normal and tangential flows at two points of a contour moving with uniform velocity $\vec{V}$. Going from point A to point B, the normal and tangential image velocities (normal flow and tangential flow, respectively) change their spatial orientation. This indeed happens from point to point due to contour curvature. The normal and tangential flows are always 90° apart.

An important property of the normal flow is that this in the *only* image velocity component that can be locally computed in the image. The tangential component can not be computed. In order to explain this, it can be assumed that the image brightness $I(\cdot,\cdot;\cdot)$ is constant when image point $P(x,y)$ at time t moves to position $P'(x',y')$ at time $\Delta t\ t' = t + \Delta t$, were $(x',y') = (x,y) + \vec{V} \cdot \Delta t$. The image velocity is considered to be constant and Δt is "small". Therefore,

$$I(x',y',t') \approx I(x,y,t) \tag{1}$$

or

$$\vec{V} \cdot \vec{\nabla} I(x,y,t) + \frac{\partial I(x,y,t)}{\partial t} \approx 0 \tag{2}$$

were' $\approx$ 'means approximate and $\nabla \equiv (\partial/\partial x, \partial/\partial y)$. Since $\vec{V} = \vec{V}_n + \vec{V}_t$ and $\vec{V}_t \cdot \vec{\nabla} = 0$, (2) reduces to:

$$\vec{V}_n \cdot \vec{\nabla} I(x,y,t) + \frac{\partial I(x,y,t)}{\partial t} \approx 0. \tag{3}$$

This means that:

$$\vec{V}_n = \hat{n}\left|\vec{V}_n\right|, \tag{4}$$

with

$$\left|\vec{V}_n\right| = \frac{\left|\dfrac{\partial I(x,y,t)}{\partial t}\right|}{\left|\nabla I(x,y,t)\right|} \tag{5}$$

$$\hat{n} \equiv \frac{\vec{\nabla} I(\cdot,\cdot;\cdot)}{\left|\vec{\nabla} I(\cdot,\cdot;\cdot)\right|} \ . \tag{6}$$

The normal flow, in distinction to the image velocity, is also a measure of local image brightness gradient orientation, and this measures implicitly includes the amount of spatial shape variability, e.g. curvature, texture orientation, etc.

Preferably, two different methods may be used to compute the normal flow in discrete images $I[i][j][k]$. One method is the 2x2x2 brightness cube method is described in B.K.P. Horn, Robot Vision, The MIT Press, Cambridge, Massachusetts, 1986. Another method is the feature based method.

In the 2x2x2 brightness cube method the spatial and temporal derivatives are approximated according to (7)-(9).

$$\partial I(x,y;t)/\partial x \approx 1/4 \times [(I[i+1][j][k]+I[i+1][j][k+1]+I[i+1][j+1][k]+I[i+1][j+1][k+1]) \\ -(I[i][j][k]+I[i][j][k+1]+I[i][j+1][k]+I[i][j+1][k+1])].$$
(7)

$$\partial I(x,y;t)/\partial y \approx 1/4 \times [(I[i][j+1][k]+I[i][j+1][k+1]+I[i+1][j+1][k]+I[i+1][j+1][k+1]) \\ -(I[i][j][k]+I[i][j][k+1]+I[i+1][j][k]+I[i+1][j][k+1])]$$
(8)

$$\partial I(x,y;t)/\partial t \approx 1/4 \times [(I[i][j][k+1]+I[i][j+1][k+1]+I[i+1][j][k+1]+I[i+1][j+1][k+1]) \\ -(I[i][j][k]+I[i][j+1][k]+I[i+1][j][k]+I[i+1][j+1][k])]$$
(9)

These discrete derivatives are computed inside the cells of a 2x2x2 brightness cube.

The feature based method is based on the following steps:

(a) Finding image points with high spatial gradients. This is implemented by: (i) smoothing the image $I(\cdot,\cdot;\cdot)$ by applying to it a binomial approximation to a Gaussian function; (ii) computing the discretised spatial image gradients $\partial\tilde{I}/\partial x \approx 1/2 \cdot (I[i+1][j][k]-I[i-1][j][k])$ and $\partial\tilde{I}/\partial y \approx 1/2 \cdot (I[i][j+1][k]-I[i][j-1][k])$; (iii) finding the subset of image points for which $|\nabla I(\cdot,\cdot;\cdot)|$ is larger than a pre-determined threshold T_{G_r}. Also, use $\partial\tilde{I}/\partial t \approx 1/2 \cdot (I[i][j][k+1]-I[i][j][k-1])$, which involves three instead of two successive frames.

(b) The normal flow is computed interactively at each feature position, e.g. point with "high" spatial gradient, by using the discrete version of (5) and (6). First, with the initial computation of the normal flow, the local image is warped according to it to refine the normal flow value. From the residual temporal derivative the residual normal flow is computed and the initial normal flow estimate is updated. This is repeated until the residual normal flow is smaller than ε (e.g. 0001).

Normal acceleration describes temporal variation of the normal flow along the normal (image brightness gradient) direction. Its importance is due to the fact that the acceleration measures how much the normal flow varies between, at least three successive frames, and thus making it enables to determine how much the space-time details *vary* between pairs of frames.

One way to define the normal acceleration is by taking the temporal derivative of (3):

$$\frac{\partial}{\partial t}\left[\vec{V}_n \cdot \vec{\nabla} I(x,y,t) + \frac{\partial I(x,y,t)}{\partial t}\right] = \vec{A}_n \cdot \vec{\nabla} I(x,y,t) + \vec{V}_n \cdot \frac{\partial}{\partial t}\vec{\nabla} I(x,y,t) + \frac{\partial^2 I(x,y,t)}{\partial^2 t} \approx 0$$

$$(10)$$

so that:

$$\vec{A}_n = \hat{n}\left|\vec{A}_n\right|, \qquad (11)$$

and

$$\left|\vec{A}_n\right| = \frac{\left|\vec{\nabla} I(x,y,t)\right| \cdot \frac{\partial^2 I(x,y,t)}{\partial^2 t} + \left|\frac{\partial I(x,y,t)}{\partial t}\right| \cdot \left|\frac{\partial \vec{\nabla} I(x,y,t)}{\partial t}\right|}{\left|\nabla I(x,y,t)\right|^2}. \qquad (12)$$

Because of the second temporal derivative in (12), it is necessary to use a minimum of three successive frames when implementing (12). Taking a $3 \times 3 \times 3$ pixels wide cube to compute the discretised versions of the derivatives in (12), it can be shown that:

$$\frac{\partial^2 I}{\partial t} \approx \frac{1}{6}[I[i][j+1][k-1]+2 \cdot I[i][j][k-1]+I[i][j-1][k-1]+I[i+1][j][k-1]+I[i-1][j][k-1]$$
$$-2 \cdot (I[i][j+1][k]+2 \cdot I[i][j][k]+I[i][j-1][k]+I[i+1][j][k]+I[i-1][j][k])$$
$$+I[i][j+1][k+1]+2 \cdot I[i][j][k+1]+I[i][j-1][k+1]+I[i+1][j][k+1]+I[i-1][j][k+1]];$$

$$(13)$$

The other discretised derivatives can be obtained to (7)-(9) on the $3 \times 3 \times 3$ cube.

The goal of computing feature statistics is to detect space-time regions were a given feature varies most – the segmentation and detection of high space-time details. This may be implemented according to the following algorithm, given two (three) successive images:

5   1. Dividing the image into non-overlapping (square or rectangular) blocks,
     2. Computing within each block a local feature set,
     3. Determining, for each block, the average of the feature set computed in 2., and
     4. Computing the variance, average variation of each feature within each block from the variance computed in 3.,
10   5. Given a threshold T_{stat}, selecting a set of blocks for which the variance computed in 4. is larger than T_{stat}.

In our implementation of the algorithm we choose square (8×8 or 16×16) blocks. This will tessellate the image into square blocks, and the remainder of it will be left
15 untessellated; in order to reduce this residual untessellated image region a rectangular tessellation could be used, but this is not so interesting because we want to align these blocks with MPEG 8x8 (DCT) or 16x16 (MC) blocks for visual artefact pre-detection. The computation of feature values within each block is implemented either at each pixel, for which $|\nabla I(\cdot,;\cdot)|$ is larger than a pre-determined threshold T, or at feature points for which
20 $|\nabla I(\cdot,;\cdot)|$ is larger than a pre-determined threshold T_{Gr}; usually $T < T_{Gr}$. The statistics exemplified in steps 4. and 5. are just an illustration. More detailed statistics could be computed. Also, specific probability distribution densities (pdf) and their statistics could be computed.

25   In order to make the computations according to the above-mentioned or related implementations more robust, a set of pre- and post-processing operations may be applied. An example of pre-processing is to convolve the input images with low-pass filters. Post-processing may include, for example, comparing neighbour blocks with respect to their statistics, e.g. feature variance.
30
        Fig. 2a shows an example of an image taken from a sequence of images. In the image two persons are watching splashing water in a fountain basin. One of the persons is partly behind the splashing water. Such an image therefore includes local parts exhibiting an example of a phenomenon expected to produce a chaotic brightness pattern, namely the
35 splashing water. Therefore, the image is taken from a moving image sequence with the potential of a high amount of local space-time details. The image has been processed according to the present invention in blocks, and for each block a variance of normal flow magnitude has been calculated as a measure representing the amount of space-time details.

In Fig. 2b the blocks of the image of Fig. 2a are shown in a grey scale indicating normal flow magnitude variance and thereby indicates the amount of local space-time details. White coloured blocks indicate regions with a high level of normal flow variance, whereas

5   dark grey blocks indicate regions with a low level of normal flow variance. As seen from Fig. 2b white blocks appear in parts of the image with splashing water and thus these local image regions are found to exhibit a large amount of local space-time details according to the processing method. The steady image regions, such as the person to the left and the fountain basin to the right, are seen to be dark grey, indicating that these regions are detected to

10   exhibit a low normal flow variance.

Fig. 3 show a flow diagram structure of a system for processing space-time details information. The system sketched in Fig. 3 can be used for different applications by using different of paths A, B and C indicated in the flow diagram. The elements of Fig. 3 are:

15

VI: Video Input
Pre-P: Pre-processing
STDE: Space-time detail estimation and detection
Post-P: Post-processing
20   VQI: Visual quality improvement
Disp: Display
St: Storage medium

Video input of Fig. 3 represents a video signal representing a sequence of images. The
25   video input may either be applied directly, such as by a wire or wireless, or as indicated in Fig. 3, the video signal may be stored on a storage medium before being processed. The storage medium may be a hard disk, a writeable CD, a DVD, computer memory etc. Input may either be a compressed video format, such as MPEG or H.26x, or it may be a non-compressed signal, i.e. a full resolution representation of the video signal. If an analog video

30   signal is input, the VI step may include an analog to digital conversion.

Pre-processing of Fig. 3 is optional. If preferred, various signal processing may be applied in order to reduce noise or other visual artefacts in the video signal before applying the space-time detection processing. This enhances the effect of the space-time detection

35   processing.

Space-time detail estimation and detection (STDE) is performed according to the above-described methods. Preferably the method includes calculation of visual normal flow and it may further include calculation of visual normal acceleration. The necessary

calculation means may be a dedicated video signal processors. Alternatively, since the amount of calculations needed with the methods according to the present invention signal processing may be implemented using signal processing power already present in the device, such as a TV set or a DVD player.

5

Post-processing may include various per block statistical methods performed on statistical results for each of the blocks of the STDE part of the system of Fig. 3. The post-processing may further include an integration in time of the statistical results for each of the blocks of the STDE step of Fig. 3. In addition, the post-processing may comprise determining

10   a pattern of temporal evolution of the per block statistics in time. This is necessary to determine which parts have a stable statistics.

Using path A of Fig. 3 the video signal is stored after detection of space-time details. Preferably, the video signal is stored together with indexing information allowing further

15   processing to be performed later.

Alternatively, visual quality improvement means may be applied before storing, i.e. path B may be used. Visual quality improvement means may be provided to the signal so as to utilise the provided information regarding local regions of images containing a large

20   amount of space-time details. For a non-compressed video signal this may be done by allocating, to blocks with space-time details, a larger data rate than would normally be allocated by standard coding schemes - for example by reducing the quantisation scale in I-frame and P-frame coding, to cope with higher levels of details. The signal may then be stored in an encoded version, however processed so as to eliminate or avoid visual artefacts.

25   The video signal may be store without encoding but provided with indexing information indicating blocks or regions with space-time details thus enabling further processing such as later encoding or using the space-time index information as a search criterion.

The last processing part of the system of Fig. 3 is a visual output, i.e. display, such as

30   on a TV screen, a computer screen etc. Alternatively, the video signal may be applied to further devices or processors before being displayed or stored.

An application (i) of the principles according to the present invention is to eliminate or at least reduce visual artefacts in a video signal, such as the artefact blockiness or temporal

35   flickering, by allocating more bits for blocks detected to exhibit space-time details. In some situations it may be preferred merely to obtain an indication of images/video regions which will contain probable visual artefacts, such as, blockiness, ringing, and mosquito "noise" for digitally (MPEG, H.26x) processed videos once encoded.

Another application (ii) is to implement a low cost motion detection indicator for field insertion in de-interlacing for TV systems that can profit from a spatial sharpness improvement. This may be especially suitable for application within low cost de-interlacers, the principles according to the invention providing a partial motion compensation

5   information.

Yet another application (iii) is to detect, segment, index and retrieve image regions detected to exhibit space-time details in long video databases. In this way it may be possible to provide a search facility that allows a quick indexing of sequences of e.g. video films that

10   contain waterfalls, ocean waves, hair/leaves/grass moving in the wind etc. Depending on which application is targeted, different processing blocks are used.

Yet, another possible application (iv) is to perform selective sharpening, i.e. to adaptively change the spatial sharpness (peaking and clipping) to highlight selected regions

15   of an image where a sharper image is desired, and to reduce the possibility of increasing the visibility of digital artefacts in regions that are de-selected.

For example, application (i) can be used in both visual quality improvements for display and storage applications. For display application path C in Fig. 5 is used. Display

20   applications may be such as high quality TV sets. Detection and segmentation of space-time details is important due to the fact that visual artefacts can be eliminated or at least reduced by an appropriate allocation of bits in response to local/regional image characteristics, such as, a customised bit-rate control per 8x8 or 16x16 image blocks. This is important relating to visual artefacts because often by just detecting may be too late to reduce their visibility or

25   effects on the visual quality of motion pictures when displayed.

In storage applications path A or path B of Fig. 5 may be used. By using path A the video signal is stored prior to performing visual quality improvement. However, using path A may include detection and segmentation of space-time details and storage of indexing of

30   regions, such as 8x8 or 16x16 pixel blocks, that contain a large amount of space-time details. In this way a long video databases (stored content) may be processed enabling a further process at a later stage. This is useful for content information that is highly detailed and for which no effective representation is known for content description. Video signals may be stored either compressed or uncompressed. By storing uncompressed data a later

35   compression can be performed taking advantage of the stored index relating to local space-time details.

By using path B video signals are stored after being properly processed with respect to increasing visual quality based on the detected local space-time details. As mentioned, the

visual quality improvement could be performed by allocating more data to blocks exhibiting a space-time details. Therefore, path B may also be used for processing large video databases. Using path B video signals can be stored compressed since a proper signal treatment has been carried out ensuring that a high visual quality regarding space-time details is obtained even by use of compression.

Among a large amount of different devices or systems, parts of devices or systems, the principles according to the invention may be applied within TV systems, such as TV sets, and DVD+RW equipment, such as DVD players or DVD recorders. The proposed methods may be applied within digital (LCD, LCoS) TV sets where new types of digital artefacts occur and/or become more visible and thus requiring a generally high video signal quality.

The principles of the present invention relating to visual quality improvement may be used also within wireless hand-held miniature devices featuring displays adapted for showing motion pictures. For example, a high visual quality of motion pictures on mobile phones with near to the eye displays can be combined with still a moderate data rate requirement. For devices with a quite poor spatial resolution the visual quality improvements according to the invention may be used to reduce the required data rate for the video signal, and still without blockiness and related visual artefacts.

In addition, the principles according to the invention may be applied within MPEG coding and decoding equipment. The methods may be applied within such encoders or decoders. Alternatively, separate video processor devices may be applied prior to existing encoders. The principles according to the invention may be applied within consumer equipment as well as within professional equipment.

In an embodiment of a video signal encoder according to the invention, a quantisation scale at the encoder side depending on space-time details information is applied. The quantisation scale is modulated by space-time details information. The smaller (larger) this scale the more (less) steps the quantizer has, and therefore more (less) spatial details is enhanced (blurred). Preferably, a video signal encoder according to the invention is capable of producing signal formats in accordance with MPEG or H.26x formats.

In a preferred embodiment, a fixed quantisation scale per macroblock q_sc is used. A modulation is applied to q_sc, wherein the modulation using information about space-time details. For each macroblock the normal flow (per pixel) and its average and variance σ_{v_n} (per macroblock) are calculated. From experiments it is known that the normal flow variance

has a histogram for which the Gamma (Erlang) function is a good fit. With this knowledge, it is possible to fit:

$$M(x) = x \times \exp(-(x-1))$$

5   (shifted Gamma (Erlang) function) to the histogram of σ_{v_n}. With this, the quantisation scale per macroblock becomes:

$$q\_sc\_m = F(\delta \times q\_sc - \lambda \times M(\sigma_{\bar{v}_n})),$$

where F() represents the operations of rounding and table look-up, and δ and λ are real
10   numbers (positive for δ and positive and negative for λ) that are adjusted according to an overall amount of bits preferred to assign per frame (video sequence).

Fig. 4 shows an example of a histogram plotted for a sequence exhibiting image parts with a high amount of space-time details. The sequence processed is the sequence of a girl
15   running in the foreground, while part of the background is the sea with water waves hitting rocks. The histogram of Fig. 4 shows a number of blocks as a function of normal flow variance. The white bars indicate flat areas, i.e. areas with a small amount of space-time details, e.g. the sky. The black bars indicate areas with a high amount of space-time details, e.g. water waves hitting the rocks. As seen from the histogram there is a good correlation
20   between space-time details and normal flow variance, since bars representing areas with small amount of space-time details are grouped towards low normal flow variance values, while bars representing high amount of space-time details are grouped towards high normal flow variance values.

25   In the foregoing, and also with regard to the accompanying claims, it will be appreciated that expressions such as "incorporate", "contain", "include", "comprise", "is" and "have" are intended to be construed non-exclusively, namely other parts or components are potentially present which have not been explicitly specified.